**RC 1:**

1) This is a generally very well written and very solid documentation of the setup and calibration procedure for flux measurements from a unique low flying aircraft. A necessary basis for any follow-on studies using observations from this particular system, but as such there is little novelty in concepts or methods or data. In highlights + associated comments in the attached document I suggest a number of mere technical/editorial improvements. In addition, a few issues require a bit more attention, some preferably in the form of an explicit discussion.

*Thanks a lot for reviewing this manuscript and providing valuable comments/ suggestions for our manuscript. We have responded to the technical/ editorial improvements in the attached .pdf document. Additionally, we have responded to the other issues that require more attention below.*

2) The integration of the Picarro trace gas analyser is poorly documented. E.g. there must a significantly long inlet tube between instrument in cabin and inlet (I assume) in wing pod. What flow rate/pump was used? How were unavoidable significant time delays dealt with (documented a bit late in the text)? Why are lag times of H2O treated differently from CO2 and CH4 if all 3 signals come from the same instrument/setup?

*Thanks for the suggestions. We have incorporated more information about the integration of the Picarro gas analyzer (tube length, response time of the gas caused by the length of the tube) in the manuscript as follows:*

*"The distance between the inlet of the tube and the gas analyzer and the five hole probe was small (< 0.5 m). The tube was ca 6 m long, had a flow rate of ca. 5.8 sL/minute and an inner diameter of ca. 0.04 m. Based on these characteristics, the transport time of the gas between the inlet tube and the G2311-f gas analyzer was ca. 0.8 seconds"*

*We have also added a table with an explanation of the lags that were calculated between the gas measurements and the vertical wind w' (new Table 4, see below). Additionally, we have written some text in the section "Temporal and Spatial Alignment Wingpod Data" to redirect to section 2.5 (and Figure 2) in the manuscript where this cross-correlation was performed - to make it easier to find for the reader. This extra sentence reads now as follows:*

*"Temporal alignment of the wingpod data and the Picarro data is performed at a later stage in the data processing using Eddy4R (see Figure 2; for details see section 2.5)."*

*The lag times of $H_2O$ are treated differently from $CO_2$ and $CH_4$, as we observed that the water vapor has more delay, when looking at the cross-correlations – $H_2O$ can adhere to the tube surface and therefore will have more delay compared to the other measured gasses. As the adherence depends on the humidity within the tube, we assume that changes in lags are possible between the legs, which is why we use leg-defined lags to correct for time offsets between $H_20$ and w.*

3) A bit more discussion and/or analysis of the differences/pros/cons of wavelet vs reynolds flux calculation would be welcome. Same for use of 2km integration windows with only 200m stepsize, which artificially reduces random errors and increases (sometime unrealistically) autocorrelations between data points. Sharp transitions in surface fluxes (e.g. lake) might be unnecessary convoluted.

*The use of 2 km integration windows with 200 m step size may indeed introduce some autocorrelation due to 90% overlapping samples, which could artificially reduce ensemble random errors:*

$$Error = \frac{\sigma}{\sqrt{N}}$$

*where:*

- *$\sigma$ is the standard deviation of the random error in individual samples.*
- *$N$ is the number of independent samples.*

*We have accounted for this by recognizing that overlapping samples are autocorrelated, and the effective sample size $N_{eff}$ is reduced accordingly, following the formula:*

$$N_{eff} = \frac{N}{1 + 2\sum_{k=1}^{K} \rho(k)}$$

*Here:*

- *$\rho(k)$ is the autocorrelation function of the sample with lag $k$.*
- *$K$ is the maximum lag where significant autocorrelation exists.*

*Using $N_{eff}$ in place of $N$ corrects the ensemble random error to reflect the increased autocorrelation between samples. The individual overlapping samples are still valuable because they preserve high spatial resolution, which is critical for capturing sharp transitions in fluxes (e.g., from land to lake) and for reducing random noise in turbulent atmospheric conditions. Additionally, wavelet-based flux calculation benefits from this approach, as it allows for multi-scale analysis and better characterization of spatial heterogeneity, compared to traditional Reynolds-averaging methods that smooth out small-scale variations.*

***Detailed Rationale for the Utility of Overlapping Samples***

1. ***Reducing Random Noise**: Even with overlapping measurements, retaining these individual samples is useful because averaging over autocorrelated samples still helps in reducing random noise. This is especially useful in turbulent atmospheric measurements, where sharp transitions in surface fluxes (e.g., from a lake to land) can cause significant noise. Reducing random noise improves the overall signal-to-noise ratio.*
2. ***Preserving Spatial Resolution**: Aggregating all samples into larger blocks (e.g., non-overlapping 2 km windows) would reduce spatial resolution and may obscure important features of the surface fluxes. For instance, in heterogeneous landscapes, preserving the high resolution (e.g., 200 m step size) helps capture localized changes in fluxes that would otherwise be lost.*
3. ***Wavelet Analysis vs Reynolds-Averaging**: In wavelet-based flux calculations, the overlapping samples allow for a finer, multi-scale analysis, where the goal is not just to compute a single flux value but to understand how fluxes vary across spatial scales. This contrasts with Reynolds-averaging, which smooths out small-scale variations. Retaining*

*overlapping samples in wavelet-based analyses ensures that transitions in fluxes (e.g., from a lake to land) are resolved.*

4) Since your turbulence probe is mounted in a wingpod, ie off the symmetry-axis of the fuselage, may be you can discuss a bit the implications this has (implicitly or explicitly) for the calibration procedures/parameterisations for the true wind vector.

*Thanks for discussion this point. Indeed, similar to Mallaun et al. (2015) and Neininger et al. (2001), the wind measurements are taken off the symmetry axis of the fuselage.*

*This would, however only be problematic if the heading of the aircraft would be very different from the horizontal orientation of the aircraft – if the angle between the aircraft's forward direction and the direction in which the aircraft is moving would be very large (see Figure RC1.1).*

*However, the actual sideslip angles are much smaller than the angle that would be required to create the situation that the probe would be shaded by the fuselage (Figure RC1.1) For example, during a measurement flight with start and landing on the 21st of August, 2019 the beta angle was max. 10° (more than a factor 5 smaller).*

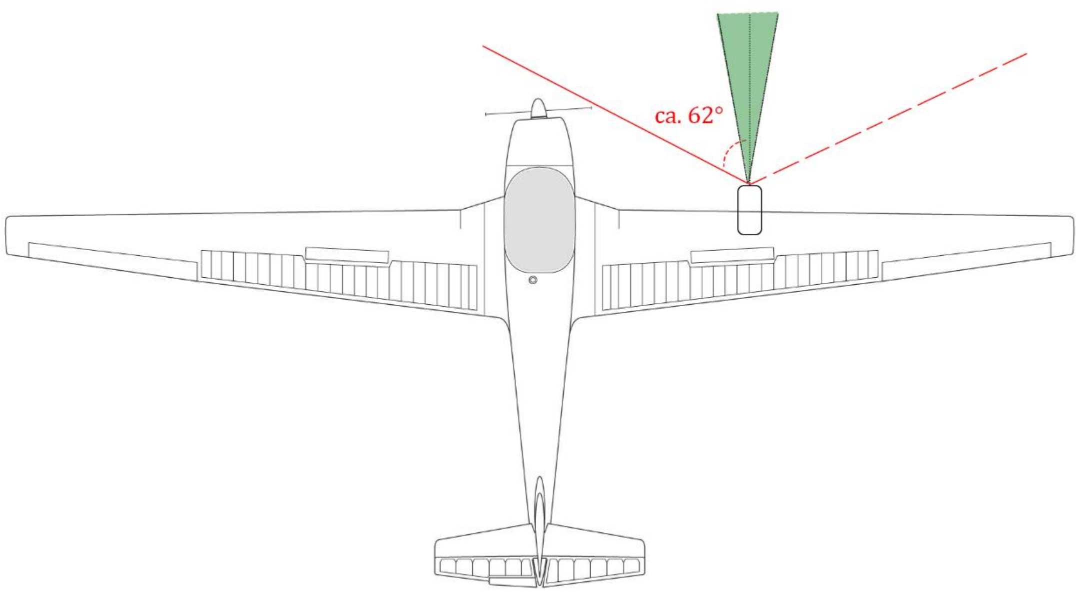


*Figure RC1.1: Position of the wingpod and the sideslip angle that would be required to cause conditions in which the probe would be shaded by the fuselage, green area indicates obtained angles from measurement flight on the 21st of August 2019 (as an example).*

5) Unlike your near-perfect wind spectra, for the other signals they are (much) less perfect. Can you *show* that this does not affect calculated fluxes as you claim, e.g. using cospectra? How might that change if you fly lower? Somewhat related, do you do any high/low freq. corrections?

*Generally, a higher level of random noise does not affect the covariance, and thus the fluxes, as the noise is not correlated with the vertical motion. This was already shown in detail by Hartmann et al. (2018).*

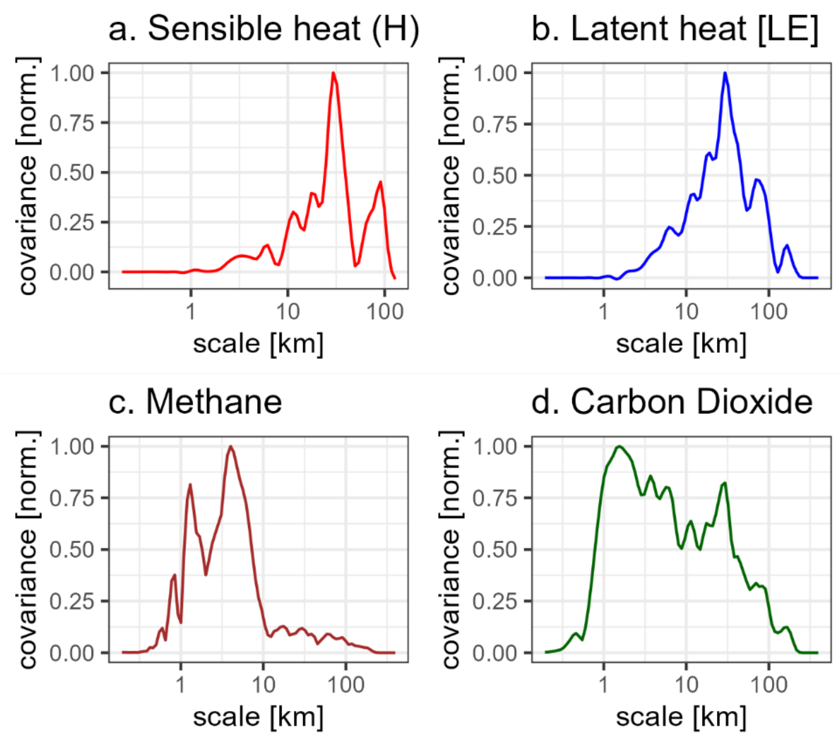


*Figure RC1.2: Averaged cospectra from flight legs 29th of August 2018 and 21st of August 2019 (in total 11 flight legs were available, for sensible heat only 6 legs were available as no fast Temperature sensor was installed in 2018).*

The cospectra that are shown in *Figure RC1.2* clearly indicate that the noise signal that was visible in the spectral plots of the temperature, and Picarro data (Figure 10b of the manuscript) is not correlated with the vertical wind and does not cause an artificial flux signal.

***References***

*Hartmann, J., Gehrmann, M., Kohnert, K., Metzger, S., and Sachs, T.: New calibration procedures for airborne turbulence measurements and accuracy of the methane fluxes during the AirMeth campaigns, Atmos. Meas. Tech., 11, 4567-4581, 10.5194/amt-11-4567-2018, 2018.*

*Mallaun, C., Giez, A., and Baumann, R.: Calibration of 3-D wind measurements on a single-engine research aircraft, Atmos. Meas. Tech., 8, 3177-3196, 10.5194/amt-8-3177-2015, 2015.*

*Neininger, B., Fuchs, W., Baeumle, M., Volz-Thomas, A., Prévôt, A., and Dommen, J.: A small aircraft for more than just ozone: Metair's' Dimona'after ten years of evolving development, 11th Symposium on Meteorological Observations and Instrumentation, Amer. Met. Soc., Boston, 123-128,*

**RC 2:**

The paper describes a new airborne system intended to measure fluxes via eddy covariance and wavelet methods. The paper describes the calibration of the system in detail and presents an example of data taken over Germany. Overall this is an excellently written paper. I think it does an good job of explaining all the different possible errors that go into making turbulent wind measurements from an aircraft. It does not spend much time on the uncertainties in the gas measurements themselves or on additional errors when these instruments are put on an aircraft. I've made some comments about this. The gas measurements are clearly not the focus of this paper and that is fine, but it might be worth pointing that out. At least that uncertainties in the gas measurements have not been evaluated.

*Thanks a lot for reviewing this manuscript and appreciate the comments and suggestions made by the reviewer. We understand the concerns of the reviewer related to the errors in the gas measurements. We have addressed these comments in detail below.*

I have some minors comments listed below

1) Line 96: insert 'and' before (2)

*Thanks a lot for the suggestion. We have adjusted the accordingly in the manuscript.*

2) Lines 120:135: It would be good to include some detail on the placement of the gas sensors and their inlets distance from the wind probe in this section. Estimated time of flight between inlet and sensor.

*We understand the importance of adding this information to the manuscript. We have incorporated this comment and have added the following text to this segment of the manuscript:*

*"The distance between the inlet of the tube and the gas analyzer and the five hole probe was small (< 0.5 m). The tube was ca 6 m long, had a flow rate of ca. 5.8 sL/minute and an inner diameter of ca. 0.04 m. Based on these characteristics, the transport time of the gas between the inlet tube and the G2311-f gas analyzer was ca. 0.8 seconds"*

3) Table 2: Have you evaluated the accuracy of the gas sensors? Short and long term drifts which particularly could affect your flux measurements. Depending on cell flush time, have you checked that the 10 Hz data are truly independent? Are the numbers here based on your own measurements using the sensors or just copied from the spec sheet from the company?

*Thanks a lot for this insightful comment. We acknowledge the importance of thoroughly evaluating the accuracy of the gas sensors, especially considering potential short- and long-term drifts that could impact flux measurements. Most of the data provided in Table 2 originate from manufacturer manuals; however, part of the information has been supplemented by on-ground measurements, which are indicated by ** in the table. We have also incorporated data from relevant peer-reviewed literature that addresses the performance of similar systems.*

*To further strengthen the evaluation, we have provided additional details about sensor setup, including response time and drift characteristics. Specifically, based on the Picarro G2311-f manual,* *the gas analyzer has a response time of at least 5 Hz, which is suitable for airborne eddy covariance measurements. Moreover, information regarding the temporal response and flushing characteristics was added to Table 1.*

*To address your specific concern regarding the independence of the 10 Hz data, we performed a spectral analysis on both the raw and processed gas concentration data. The analysis confirms that the 10 Hz data remain independent due to the sufficiently fast cell flush time, which allows for rapid response to atmospheric changes. This is supported by prior studies (Peltola et al., 2014; Yang et al., 2016), which validate the performance of cavity-based analyzers in similar flux measurement setups. These studies report that the sensor systems are capable of high temporal resolution flux measurements with minimal drift and noise, and their findings align with the specifications and operational characteristics of our setup.*

*Finally, we have taken steps to monitor potential short- and long-term drifts in the sensor calibration by periodically calibrating the instrument against known reference gases and performing diagnostic checks during each measurement campaign. These procedures minimize the potential impact of sensor drift on the final flux calculations.*

4) Line 169: data 'were' merged (not was)

*Thanks! We have adjusted the accordingly in the manuscript.*

5) Lines 224:226 You might want to add something about the lag between the Picarro sensors here or at least mention that it'll be discussed later as this seems a natural section to the reader.

*Thank you for this important suggestion. The lag between the Picarro gas analyzers and the wind measurements is indeed the focus of our approach and is fully described in Section 2.5. As highlighted in Metzger et al. (2017), we determined lag times for each flight leg using a high-pass filtered cross-correlation between the vertical wind velocity (w') and gas concentrations ($H_2O'$, $CO_2'$, and $CH_4'$), following the method proposed by Hartmann et al. (2018).*

*Recent developments, such as the work by Vitale et al. (2024), have further popularized the use of high-pass filtering for lag detection in large eddy-covariance station networks like ICOS in Europe and NEON in the US. This method, which is gaining widespread endorsement by organizations such as FLUXNET, particularly for processing fluxes of low-concentration species (e.g., $CH_4$ and $N_2O$), offers a robust approach for temporal alignment. It is notable that the same Hartmann et al. (2018) method we applied is also being used in these networks, underscoring the robustness of this approach for high-quality flux measurements across different platforms.*

*In our manuscript, we employed this cross-correlation technique to account for the small but significant temporal offsets between gas concentration measurements and vertical wind components. This is critical for ensuring accurate flux calculations, especially for species like $CH_4$ and $H_2O$, where time delays due to transport through the measurement system or sensor response time can introduce uncertainties if not properly accounted for.*

*We will add a brief reference to this procedure in the earlier sections of the manuscript (e.g., in lines 224–226), directing the reader to Section 2.5, where the full details of our lag-correction methodology are presented. The new sentence will read as follows:*

*"The lag between the gas analyzers and the wind measurements is corrected using a high-pass filtered cross-correlation technique as detailed in Section 2.5 (Metzger et al., 2017; Hartmann et al., 2018)."*

*This inclusion should provide clearer guidance to the reader on how we handle lag corrections, while also positioning our methodology within the context of current best practices in the field.*

6) Line 406: Could you provide some rational to using 200 m and 2000m for the distances. Did you use Ogive analysis or some other method to determine the length needed to sum over relevant frequencies. 200 m seems especially short.

*The step size and window length used in our flux calculation were chosen based on previous work by (Metzger et al., 2012; Metzger et al., 2013), taking into account the altitude of the aircraft, atmospheric mixing, the characteristic length scales and resolution of surface features.*

*The window size for flux calculation, set to 2000 m, is not arbitrary but is designed to balance the trade-off between random error (which decreases with larger window sizes) and resolution (which increases with smaller windows). As shown in Metzger et al. (2013), longer windows reduce random flux error due to the inverse proportionality between random error and the square root of the averaging length (Lenschow and Stankov, 1986).*

*Importantly, the 2000 m window length does not limit the inclusion of low-frequency transport scales into the wavelet fluxes. This is because the wavelet transform orthogonally decomposes time (or position along the flight path) and transport scales, ensuring that all atmospheric transport scales contribute fully to the fluxes at each position along the flight path. The 2000 m window acts more as a low-pass filter to reduce noise, improving the signal-to-noise ratio of the flux measurements. The step size of 200 m, while resulting in 90% overlapping samples, ensures fine spatial resolution and reduces the influence of noise while fully accounting for the degree of autocorrelation in aggregate uncertainty estimates.*

*In terms of determining the appropriate window and step sizes, Metzger et al. (2013) provides a detailed rationale for the wavelet-based method, particularly in Section 3.1. This approach accounts for the need to resolve flux changes on a spatial scale that corresponds to the characteristic heterogeneity of the surface and atmospheric blending effects that limit spatial flux resolution as a function of measurement height. For example, from an aircraft measurement height of less than 100 m the upwind distance where 80% of the flux contributions are included is on the order of 1000 m, which makes a 1000 m window for flux calculations physically meaningful. With a measurement height of 200 m in the present study, a longer window of 2000 m is used here.*

*For additional details on the rationale behind the selected window and step sizes, readers can refer to Metzger et al. (2013), which discusses the balance between resolution and error in flux calculations over heterogeneous landscapes.*

*This rationale will be incorporated into the manuscript.*

7) Line 620:621: Did you try doing a null experiment to check that the gas measurements are uncorrelated with the wind. In general I didn't see any description of how you calibrated the gas sensors. While in theory the noise should be uncorrelated, there may be changes in alignment, valves (leading to pressure changes), etc. that may correlate with vertical turbulence but may appear to be noise. Running calibration gas through the system and seeing that you get something close to zero flux would show that there really was not correlation between the gas measurement 'noise' and atmospheric turbulence.

*Thanks for this suggestion. We did not perform a null experiment to check that the noise was uncorrelated. Nonetheless, the random shuffling method according to Billesbach (2011), which we used for defining the detection limit should probably have revealed such issues. As mentioned also in comment 5 of reviewer 4, a higher level of random noise does generally not affect the covariance, and thus the fluxes, as the noise is not correlated with the vertical motion. This was already shown in detail by Hartmann et al., 2018.*

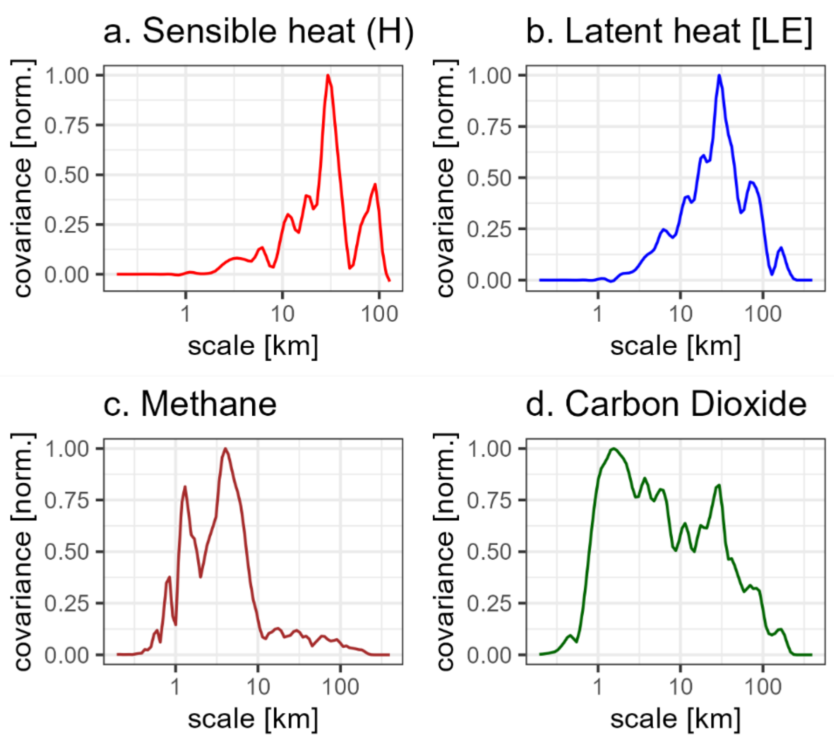


*Figure RC2.1: Averaged cospectra from flight legs 29th of August 2018 and 21st of August 2019 (in total 11 flight legs were available, for sensible heat only 6 legs were available as no fast Temperature sensor was installed in 2018).*

The cospectra that are shown in *Figure RC2.1* clearly indicate that the noise signal that was visible in the spectral plots of the temperature, and Picarro data (Figure 10b of the manuscript) is not correlated with the vertical wind and does not cause an artificial flux signal.

8) Line 680: Do you really mean uncertainty here or variability. You are listing uncertainty with an uncertainty. And given the range of uncertainty, it may be more than 100% of the flux?

*Thanks for this question/ suggestion. We actually mean variability instead of uncertainty. We have adjusted this in the manuscript accordingly:*

*"Although part of the differences in fluxes might be assigned to differences in footprints, it does give an indication of the uncertainty of the obtained fluxes. Based on the repeated flight legs, the variability in $CH_4$ fluxes was 86.2 ± 57.7 %, the variability in $CO_2$ fluxes was 32.9 ± 12.9 %, and the variability in latent heat fluxes was 36.6 ± 13.0 % per 200 m segment".*

9) Line 681: You have clearly used twice in a row.

*Thanks for this suggestion, we have removed the second "clearly" from the text. Now this line reads as follows:*

*"Clearly, Fig. 12 shows that even when we consider these uncertainties, general trends in energy and matter fluxes can still be clearly identified."*

***References:***


*Billesbach, D. P.: Estimating uncertainties in individual eddy covariance flux measurements: A comparison of methods and a proposed new method, Agricultural and Forest Meteorology, 151, 394-405, https://doi.org/10.1016/j.agrformet.2010.12.001, 2011.*

*Hartmann, J., Gehrmann, M., Kohnert, K., Metzger, S., and Sachs, T.: New calibration procedures for airborne turbulence measurements and accuracy of the methane fluxes during the AirMeth campaigns, Atmos. Meas. Tech., 11, 4567-4581, 10.5194/amt-11-4567-2018, 2018.*

*Lenschow, D. H. and Stankov, B. B.: Length Scales in the Convective Boundary Layer, Journal of Atmospheric Sciences, 43, 1198-1209, https://doi.org/10.1175/1520-0469(1986)043<1198:LSITCB>2.0.CO;2, 1986.*

*Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, Atmos. Meas. Tech., 5, 1699-1717, 10.5194/amt-5-1699-2012, 2012.*

*Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., and Desai, A. R.: eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5, Geosci. Model Dev., 10, 3189-3206, 10.5194/gmd-10-3189-2017, 2017.*

*Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Trancón y Widemann, B., Neidl, F., Schäfer, K., Wieneke, S., Zheng, X. H., Schmid, H. P., and Foken, T.: Spatially explicit regionalization of airborne flux measurements using environmental response functions, Biogeosciences, 10, 2193-2217, 10.5194/bg-10-2193-2013, 2013.*

*Peltola, O., Hensen, A., Helfter, C., Belelli Marchesini, L., Bosveld, F., Van den Bulk, W., Elbers, J., Haapanala, S., Holst, J., and Laurila, T.: Evaluating the performance of commonly used gas analysers for methane eddy covariance flux measurements: the InGOS inter-comparison field experiment, Biogeosciences, 11, 3163-3186, 2014.*

*Vitale, D., Fratini, G., Helfter, C., Hortnagl, L., Kohonen, K.-M., Mammarella, I., Nemitz, E., Nicolini, G., Rebmann, C., Sabbatini, S., and Papale, D.: A pre-whitening with block-bootstrap cross-correlation procedure for temporal alignment of data sampled by eddy covariance systems, Environmental and Ecological Statistics, 31, 219-244, 10.1007/s10651-024-00615-9, 2024.*

*Yang, M., Prytherch, J., Kozlova, E., Yelland, M. J., Parenkat Mony, D., and Bell, T. G.: Comparison of two closed-path cavity-based spectrometers for measuring air–water CO2 and CH4 fluxes by eddy covariance, Atmos. Meas. Tech., 9, 5509-5522, 10.5194/amt-9-5509-2016, 2016.*